nature human behaviour

ARTICLES
https://doi.org/10.1038/s41562-021-01066-z

Check for updates

The measurement of partisan sorting for 180 million voters

Jacob R. Brown[1,2] ✉ and Ryan D. Enos[1,2] ✉

Segregation across social groups is an enduring feature of nearly all human societies and is associated with numerous social maladies. In many countries, reports of growing geographic political polarization raise concerns about the stability of democratic governance. Here, using advances in spatial data computation, we measure individual partisan segregation by calculating the local residential segregation of every registered voter in the United States, creating a spatially weighted measure for more than 180 million individuals. With these data, we present evidence of extensive partisan segregation in the country. A large proportion of voters live with virtually no exposure to voters from the other party in their residential environment. Such high levels of partisan isolation can be found across a range of places and densities and are distinct from racial and ethnic segregation. Moreover, Democrats and Republicans living in the same city, or even the same neighbourhood, are segregated by party.

Segregation between human social groups is associated with a range of profoundly negative outcomes, including intergroup conflict, prejudice, inefficient resource allocation, poor democratic governance and other socially deleterious effects[1–4]. Segregation is also implicated in topics of intense interest across the social sciences, including interpersonal contact and intergroup relations[5,6], the bridging nature of social networks[7–11], poverty[1,12] and political representation[13,14]. Drawing on these associations and using aggregate data, popular and scholarly accounts of politics in the United States—and, increasingly, other Western democracies—describe stark partisan segregation, with members of different political parties living separate lives, resulting in partisan rancour and threatening the functions of the democracy[14–18]. Yet, despite the association between segregation and important outcomes, and the claims of increasing partisan segregation, the measurement of segregation among partisans, as with the measurement of segregation for most social groups, is severely limited: researchers must usually rely on data aggregations that do not include the actual locations of individuals, and thus measurements are limited to summaries across large geographical areas, and the experience of individual exposure across groups is masked.

In this article, using data on the exact residential address of every registered voter in the United States and harnessing advances in spatial data computation, we measure the local partisan segregation for each of these voters, creating a spatially weighted measure of cross-partisan exposure for more than 180 million individuals. These data create a large-scale measure of individual spatial segregation and yield evidence of the extent of partisan segregation in the United States, allowing us to examine the degree to which individuals are sorted by partisanship with respect to individual neighbours and within small geographic units, such as cities or neighbourhoods.

A large proportion of US voters live with very low levels of residential exposure to neighbours from the other party. The most extreme political isolation is found among Democrats living in high-density urban areas, with the most isolated 10% of Democrats in the United States expected to have 93% or more of encounters in their residential environment with other Democrats. Similarly high levels of partisan isolation are also present for Republicans living in rural areas. Such high levels of segregation may imply little exposure to competing political ideas from neighbours. In general, for voters of both parties, high levels of segregation can be found across a range of places and densities, and are distinct from, and sometimes in tension with, racial segregation. Moreover, even when Democrats and Republicans live in the same city—or even the same neighbourhood—they are residentially sorted by political party.

These high levels of partisan isolation have several important implications. In the United States, political party affiliation is considered a social identity, analogous to race or religion[19], and is a powerful predictor of a range of attitudes and behaviours[20], including behaviours outside of the explicitly political realm[21,22]. Because partisanship is correlated with political ideology and other attitudes and behaviours, the extent of a voter's partisan isolation is likely to affect their exposure to individuals different from themselves and to competing sociopolitical viewpoints, thus affecting a range of important outcomes. Cross-group exposure can be consequential for the shaping of intergroup attitudes and behaviours[6], including the prejudicial attitudes that are levelled across parties in the United States[23].

Isolated partisan environments may also affect behaviour through channels other than (a lack of) interpersonal contact: indeed, human behaviour can be shaped by low-level environmental cues[24,25], such as the norms displayed by neighbours, and randomized controlled trials have shown that political messaging from neighbours, such as the posting of yard signs, has a persuasive effect on voting behaviour[26].

Isolation may also contribute to the increasing ideological extremity on both the mass and elite levels in the United States[27]: in the marketplace for political ideas, exposure to out-partisans may enable the transmission of competing views[28] that can reduce extremism[29]. Furthermore, the extremity of political views is correlated with political participation[30] and participation is correlated with influence[31], raising the potential that the most isolated partisans are the most politically influential.

Of course, even with residential segregation, cross-group exposure may happen in other environments (for example, S. Athey et al., unpublished manuscript). But residential segregation has been

[1]Institute for Quantitative Social Science, Harvard University, Cambridge, MA, USA. [2]Department of Government, Harvard University, Cambridge, MA, USA. ✉e-mail: jrbrown@g.harvard.edu; renos@gov.harvard.edu

shown to shape attitudes and behaviour even when accounting for interpersonal cross-group contact and even with widespread access to the Internet and other communication technology[4]. Segregation has also been shown to reduce cooperation for shared benefit across groups[32] and the division of partisans in geographic space is associated with levels of trust in government and anti-system attitudes[33]. Once in place, the contribution of segregation to these behaviours can become self-reinforcing if partisans avoid living in areas where they would be a minority; thus, even initially small levels of clustering could drive extreme segregation[34], further separating partisans and reinforcing behavioural and attitudinal separation[1].

Furthermore, while the scholarly focus on racial segregation has often focused on interpersonal contact and other psychological mechanisms, the segregation of partisans probably shapes social outcomes through many other channels. In particular, segregation may also affect campaign and other elite behaviour by allowing politicians to narrow the ideological appeal of their message[35,36], thereby reducing the potential for voters to be exposed to cross-cutting appeals and allowing politicians to avoid moderating their messages. Further, the clustering of partisans in space can lead to bias in geographically based electoral districts, threatening equitable representation in government[14].

Results

To calculate partisan segregation, we used data containing information on every one of the 180,735,645 registered voters in the United States as of June 2018. Until very recently, when some states introduced automatic registration of citizens, voter registration was voluntary in nearly every state. The data include about 80% of the voting-eligible population, which is about 92% of the approximately 250 million people in the United States over 18 years of age, and does not include non-citizens and those who are not allowed to vote in certain states because of felony convictions or other reasons. With these data, we have the social group membership and exact address of almost 75% of the adult population of the United States.

When a person registers to vote, they provide a home address and, in most states, declare affiliation with a political party. We use these data to construct measures of segregation, leveraging advances in geographic data science; using Geohash techniques that store latitude and longitude coordinates as strings rather than locations, we can efficiently measure the spatial relationships between large numbers of individuals[37]. For each of $n = 180{,}660{,}202$ geocoded individuals, we measure the distance to their $k = 1{,}000$ nearest neighbours as defined by the closest geodesic distances from the registered voters' residences, creating a distance measure for $n \times k$ (over 180 billion) dyadic relationships. Thus, for every voter, we identify how near they live to each of their 1,000 nearest neighbours and combine this information with data on their neighbours' partisanship to construct individual-level measures of partisan exposure and isolation. We decided on $k = 1{,}000$ by testing a random sample of 1,000 voters with $k = 50{,}000$, and found that this provided little additional information (Supplementary Information).

With these data, we are able to implement spatially weighted measures of segregation[38] at a large scale. Measures of segregation typically available to researchers are not spatially weighted and do not use individual data, and are therefore based on the composition of groups in a chosen geographic unit. These measures usually have to make the unrealistic assumption of common geographic context for all individuals within the chosen unit, and are potentially subject to common problems of aggregate measurement, including the modifiable-areal-unit problem and problems of scale, which mean that measures of segregation can be extremely sensitive to researcher choices of geographic unit.

A standard measure of segregation is exposure, which captures the extent to which members of one group live around members of another group (or their own group, in the case of isolation)[39]. This and other standard measures of segregation are aspatial, making the stringent assumption that where individuals live in relation to each other within the geographic unit has no bearing on exposure. This assumption creates a checkerboard problem, wherein the measures are unable to distinguish between different spatial distributions of individuals even if the likely exposure was very different across the distributions[40]. In Fig. 1, for example, using the standard measure of exposure the levels of segregation on the top left and top right would be the same, despite the starkly different spatial relationships across groups. When measuring partisan segregation, because partisans are known to cluster in certain types of places (for example, Democrats cluster in the densest parts of cities[41]), not accounting for the spatial relationship of voters may obscure important variation between individuals and may lead to misleading inferences (Fig. 1). Indeed, using the measures we introduce below, we calculate that for half of voters, not accounting for distance distorts exposure by 22% or more; for 25% of voters, the distortion is 40% or more; and, in some extreme cases, the distortion exceeds 100% (Extended Data Figs. 1–3).

By measuring the distance between individuals, our measures are not subject to these issues. Our primary measure is a weighted average of exposure[38] in which the proportion of people associated with each party among an individual's 1,000 nearest neighbours is weighted by the inverse of the distance in metres from each neighbour ($Weight_k = \frac{1}{d+1}$, where d is the distance of neighbor k from voter i). Weighting by distance gives greater emphasis to an individual's closest neighbours, so the partisanship of a next-door neighbour is more important when describing partisan exposure than that of a more distant neighbour (Fig. 1). These individual-level measures of spatial exposure also enable us to flexibly explore segregation at any level, from segregation from one's most immediate neighbours up to any arbitrarily large level of geography (Fig. 2).

In measuring partisan segregation, we confront the challenge of how to account for voters who cannot or do not explicitly declare membership in a party. Partisanship is recorded at the time voters register in 30 states and in DC. If we were to measure segregation only in these states, we would miss large sections of the United States. Furthermore, in all states, some voters choose not to register with one of the two major political parties, even if they have the option of doing so. If we were to measure segregation only among voters officially registering as Democrats or Republicans, we would clearly misrepresent the levels of isolation or exposure to voters with similar or different political ideologies; extensive evidence shows that all but a small proportion of officially independent and minor-party voters have stable preferences for one of the major parties and ideological orientations indistinguishable from those of major party members[42–45].

Thus, before constructing measures of segregation, we impute partisanship for voters who are not registered as Democrats or Republicans. Such imputation techniques, relying on similar information, are commonly used by political campaigns. We impute using a three-step process in which we first code a voter as a Democrat or Republican on the basis of the last partisan primary in which they cast a ballot; for example, if a voter votes in a primary election to select Democratic candidates for office, we impute that voter as a Democrat. Next, we classify voters registered to a third party with a clear left or right ideological leaning as Democrats (left) or Republicans (right). Using these updated counts, we then impute partisanship for the remaining independents through a Bayesian process using priors constructed from 2016 precinct-level presidential vote share and individual-level demographic characteristics. This is a process similar to imputation methods for race that have been used successfully in academic research[46,47] and by political campaigns[36] (see Extended Data Fig. 4 for sensitivity of results to the imputation and Supplementary Information for details of the imputation process and summary statistics).

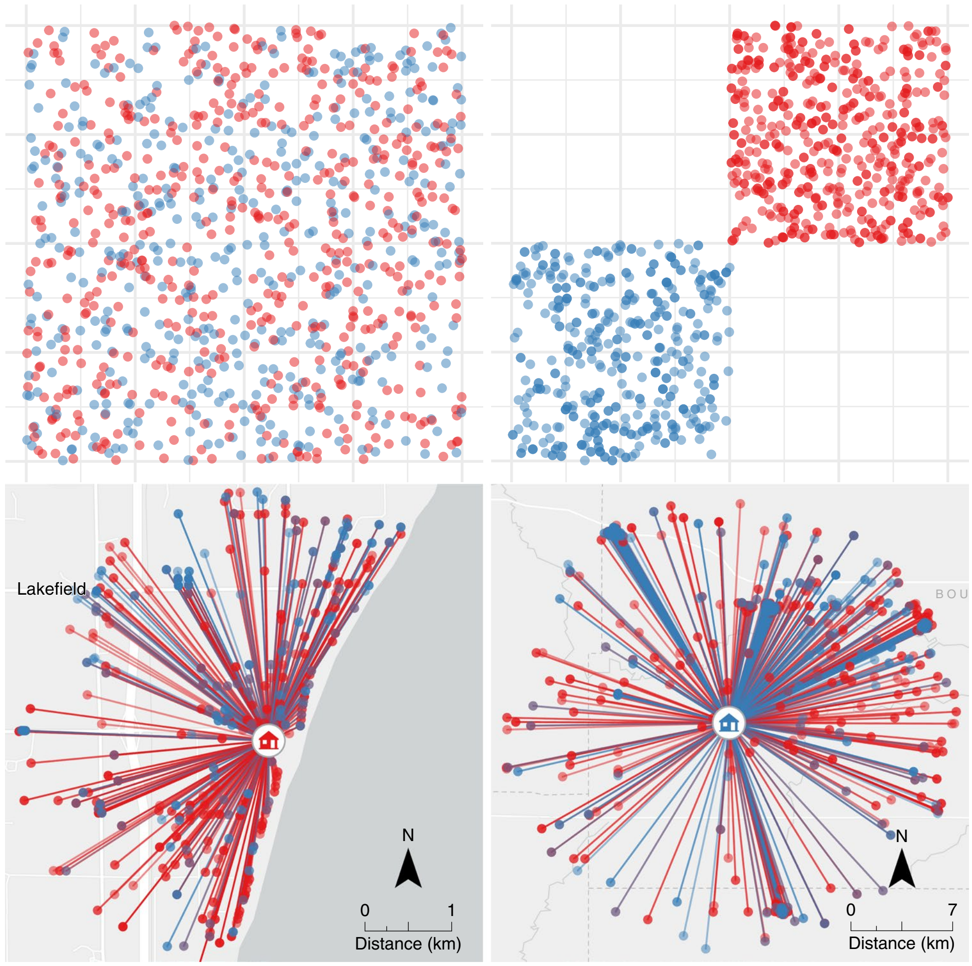


Fig. 1 | Spatial and aspatial measures of segregation. Using aspatial measures of segregation, the average exposure of red and blue individuals ($n=1,000$) in the top left and and top right maps are the same. Spatial weighting captures the apparent much higher segregation in the top right compared with the top left. Drawing on real data, a Republican in a suburb of Milwaukee, WI (red house, bottom left), has an unweighted exposure to Democrats (blue) of 0.36 among the $k=1{,}000$ nearest neighbours, but because other Republicans (red) are also clustered along the lake shore, many of whom are less than 1 km away, the spatial exposure to Democrats is 0.15. A Democrat in rural southeastern KS (blue house, bottom right) has an unweighted exposure to Republicans of 0.64 among the $k=1,000$ nearest neighbours, but because other Democrats in the area are clustered in the centres of small towns more than 7 km away, the voter is isolated from other Democrats and their spatial exposure to Republicans is 0.98. Base maps (bottom row): ESRI.

Then, using contact information from the voter file, we validate these imputations in a survey of 12,221 voters, finding that we impute voters' partisanship with an accuracy of 77%. This is near the maximum accuracy that may be possible given the instability observed in previous surveys of self-reported partisanship[19], indicating a high degree of accuracy for our method of imputation (Extended Data Fig. 5). We also show that voters with imputed partisanship are nearly ideologically identical to non-imputed voters, both in states with partisan registration and in those without, indicating that our imputation of those not registered as partisans accurately models the exposure to competing political ideas experienced when sharing residential environments with registered partisans (Supplementary Information).

These imputations yield a probabilistic score of being a partisan. When calculating isolation for voter i, we weight the contribution of each neighbour k by these posterior partisan probabilities, which will, on average, recover the probability of encountering members of the other party in their local environment. When making aggregate summaries of Democrat and Republican exposure (isolation), we also weight central tendencies and distributions by these probabilities.

With these imputations, we may be capturing the latent partisan tendencies of geographic areas, similar to the political science concept of the 'normal vote'[48]. In the Supplementary Information, we show that our measure of partisanship is highly correlated ($r=0.92$) with standard measures of the normal vote[49] at the county level.

Although our main analysis treats partisanship as a probability of being a Democrat or Republican, rather than a binary classification, the assignment of voters to discrete partisan categories may not capture an individual's latent strength of identification with these categories. This limitation is shared by standard measures of other social identities and segregation, including those measuring race, where discrete categories may not capture the strength of psychological attachment to those categories. While this is partially driven by technology—for example, it is not possible to survey every person in a large area to learn the strength of their partisan or racial identification—it is consistent with the psychology of group attachment that views categories as discrete (for example, in ref. [50]) and makes our measure consistent with other measures of segregation that rely on discrete categorizations.

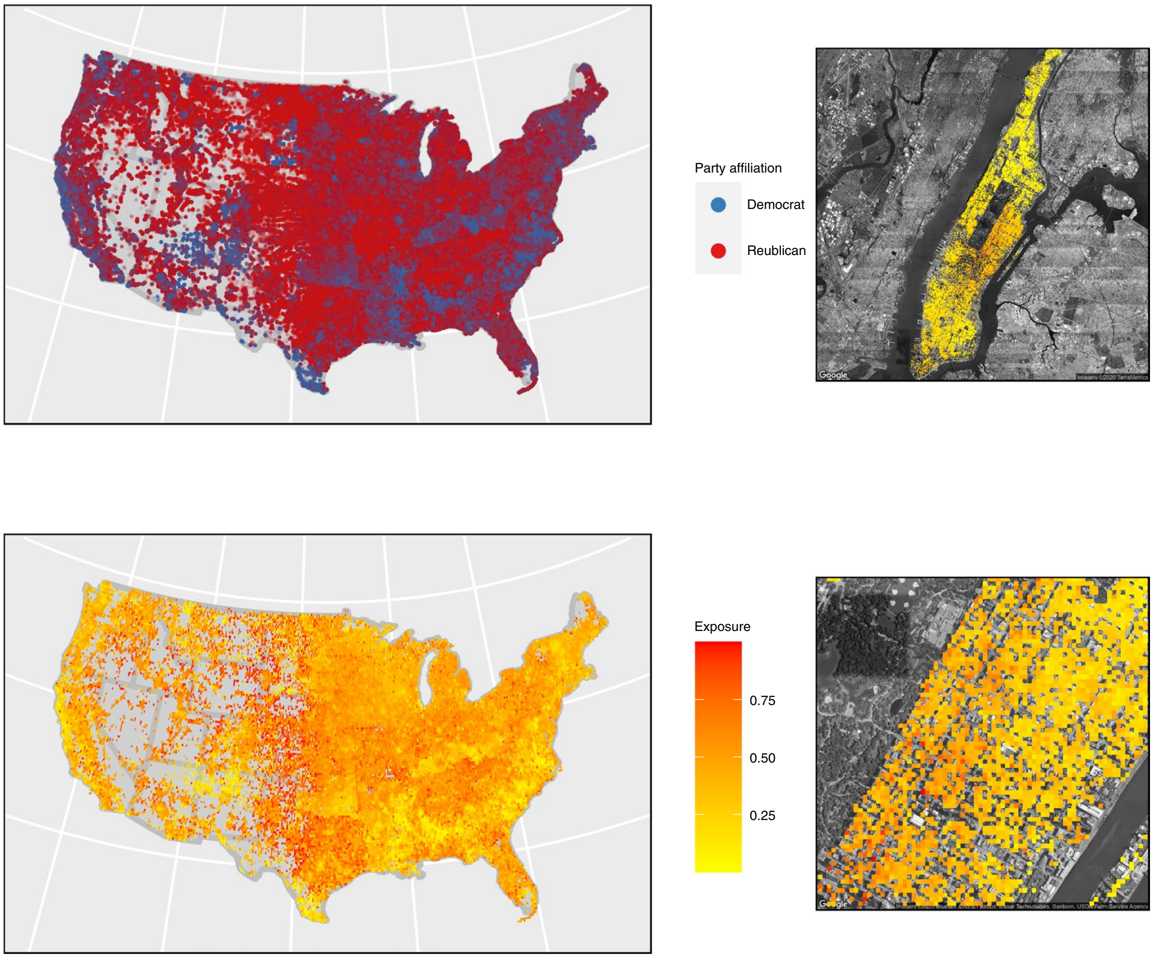


Fig. 2 | Measuring spatial exposure across increasingly small geographies. The exact residential location of every Democrat and Republican in the United States (n = 180,660,202, top left) can be used to measure each Democrat's spatial exposure to Republicans, and this can be averaged across arbitrarily small grid cells for display purposes (1,000 × 1,000 grid, bottom left). Exposure can be examined across any resolution: markedly different residential exposure to Republicans can be seen in Manhattan, NY (500 × 500 grid, top right), with Democrats on the northern and southern extremes of the island having almost no residential exposure to Republicans, whereas Democrats on the Upper East Side (the neighborhood immediately to the right of the lower section of Central Park, which is the long rectangle with no voters in it located in the center of the island) have exposure as high as 0.5 due to the clustering of Republicans in this area. A magnified view of the Upper East Side of Manhattan (75 x 75 grid, bottom right) shows the clustering of Republicans along Central Park and thus Democrats' decreasing exposure to Republicans moving towards the northeast. Map data (righthand figures): Google, TerraMetrics.

Spatially weighted measures of partisan isolation and exposure are defined as:

$$\text{Spatial isolation}_i = \frac{\sum k = 1^{1,000} \frac{1}{(dk+c)^a} \mathbb{P}(pk = pi)}{\sum k = 1^{1,000} \frac{1}{(d_k+c)^a}}$$

$$\text{Spatial exposure}_i = \frac{\sum_{k=1}^{1,000} \frac{1}{(d_k+c)^a} \mathbb{P}(p_k = q_i)}{\sum_{k=1}^{1,000} \frac{1}{(d_k+c)^a}}$$

where p_i is the partisan identification of voter i, q_i is the opposite partisan identification (if Democrat, Republican, and if Republican, Democrat) of voter i, p_k is the partisan identification of neighbour k, $\mathbb{P}(p_k = p_i)$ is the posterior probability that neighbour k has the same partisanship as neighbour i, d_k is the distance in metres neighbour k lives from voter i, c is a constant adjustment made so that when $d_k = 0$ the expression is not undefined, and a is an exponent to which we raise the denominator of the distance weight to control how much weight is given to proximity in the measure. In the main analysis, we set $c = 1$ and $a = 1$. Setting c higher would decrease the weight given to the smallest distances, and setting a higher would give greater weight to distance in general, placing even more emphasis on the closest neighbours and increasing the intensity of segregation (we show results with other weighting schemes in the Supplementary Information).

Spatial exposure and isolation represent a person's residential partisan experience. Of course, there can be partisan exposure outside of the residential context, for which we cannot account, and other variables within a voter's residential context, such as the density of non-voters, can influence the likelihood of interaction with partisan neighbours. Spatial exposure and isolation represent the potential exposure that come from the spatial arrangement of partisans: the likelihood, all else being equal, of encountering a neighbour of a given party in one's residential life. Exposure ranges from 0 to 1, with 0 being no exposure to the other party and 1 being only exposure to the other party. Isolation of 1 is perfect isolation, encountering only one's own party, and 0 is encountering only the other party. Exposure of 0.01 would mean that we expect, all else being equal, only 1 out of 100 interactions in a voter's residential context to be with a person from the other party. Exposure to the

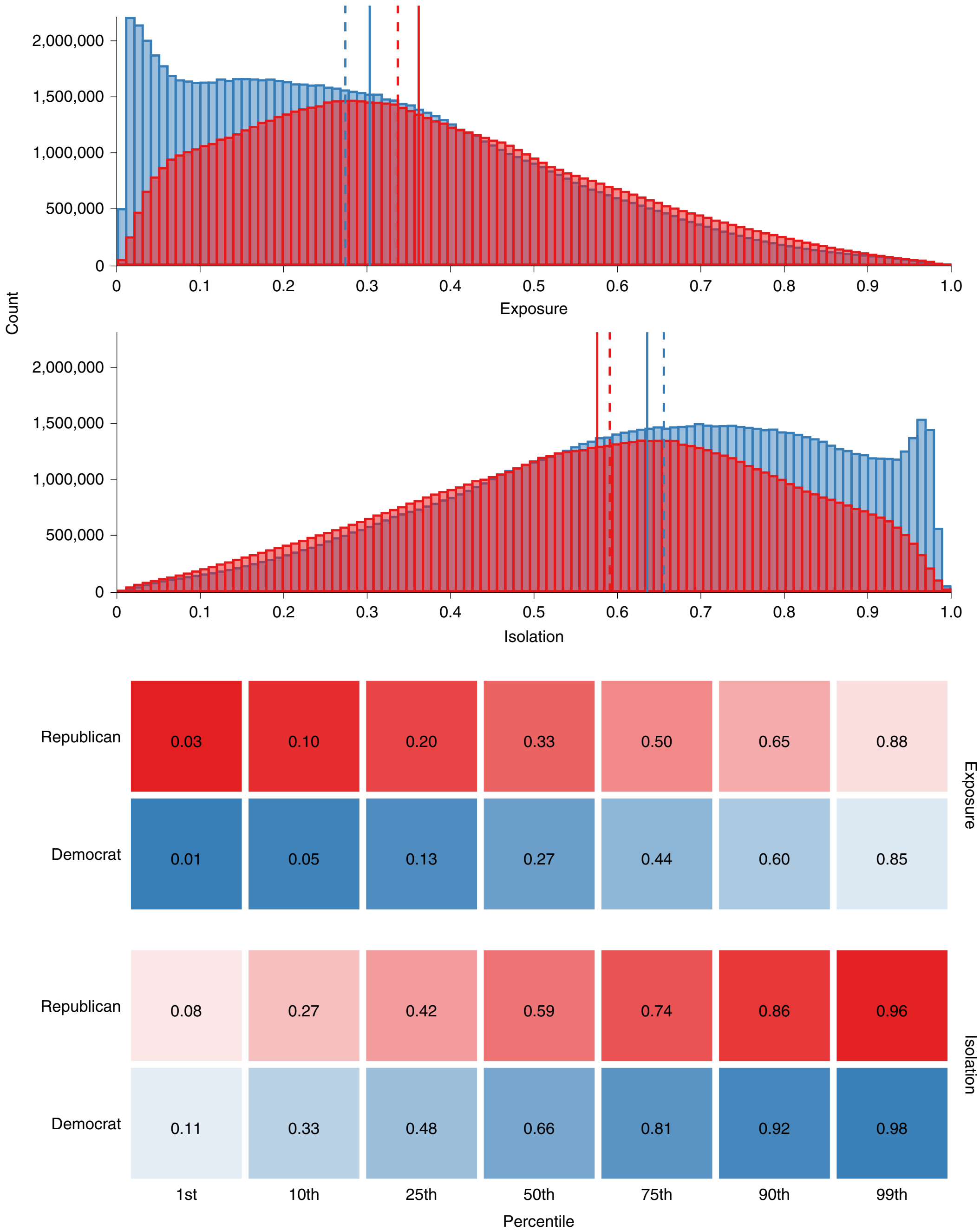


Fig. 3 | Nationwide distribution of partisan spatial isolation and exposure. Democrats are shown in blue and Republicans are in red. Solid vertical lines represent mean values and dashed lines represent median values. Coloured cells present spatially weighted proportions of out-party (exposure) or in-party (isolation) neighbours across percentiles. The distributions are weighted by the posterior partisan probabilities (n = 180,660,202 voters).

other party is not necessarily the inverse of isolation, because a small proportion of voters are true independents, neither registering with nor otherwise aligning with either party. Thus, voters not perfectly isolated from their own party might still not have any exposure to voters from the other party.

Average exposure. The average Democrat's exposure to Republicans is 0.30 and the average Republican's exposure to Democrats is 0.36. Given that the nationwide proportions of Republicans and Democrats are 43% and 51%, respectively, this represents substantially lower cross-party exposure than would be expected if partisans were not segregated into different residential environments. For both the median Democrat and the median Republican, only about 3 in 10 of their interactions in their residential environment will, on average, be with a member of the other party, all else being equal. Moreover, the national distribution (Fig. 3) shows a large portion of voters living in extreme isolation; nationwide, 10% of Democrats live with virtually no exposure to out-partisan neighbours (exposure $<$ 0.05) and a majority of Democrats and Republicans live with isolation levels well above the threshold of 0.60 commonly used to describe high isolation in the context of race in municipal areas[1].

Democrats and Republicans are segregated in most types of places. Partisan segregation is not distributed evenly across the United States and thus the level of residential interaction across party lines will vary depending on where a voter lives. The flexibility

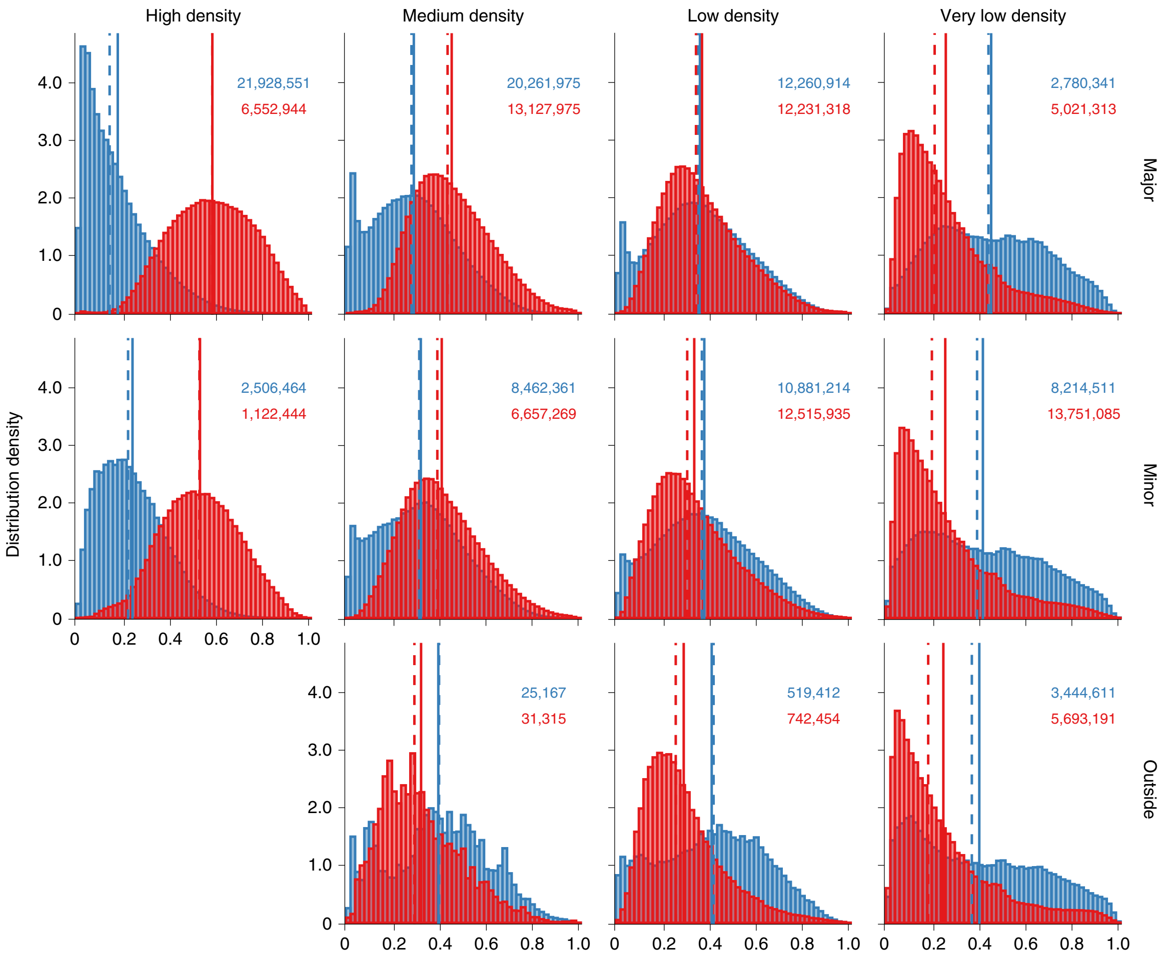


Fig. 4 | Isolation by density and urban area. Histograms shows individual-level spatial partisan exposure subset by the population density (high, medium, low or very low) of the voter's census tract and the type of urban area (major, minor or outside metropolitan area) in which they live, separately for Democrats (blue) and Republicans (red). Solid vertical lines represent mean values and dashed lines represent median values. Distributions are weighted by posterior partisan probabilities. Numbers in blue and red represent estimates of number of Democrats and Republicans, respectively, in the urban area–density subset.

of individually measured segregation allows us to describe segregation across places defined in any way and thus we can observe how partisans are likely to interact across the suburbs and central cities of both large and small urban areas, as well as rural locations.

In major urban areas (core-based statistical areas (CBSAs) with more than 1 million residents, the largest of which is the New York–Newark–Jersey City area and the smallest of which is the Tuscon, AZ area) Democrat exposure to Republicans is extremely low, especially in the dense urban cores (Fig. 4). Notably, a large plurality of Democrat voters live in these areas and the very low levels of exposure extend even to the medium-density suburbs of these major areas and to minor urban areas (those with fewer than 1 million residents, the largest of which is Honolulu).

It is only when reaching areas where relatively few voters are located—the very-low-density areas on the fringe of or entirely outside urban areas—that Republican exposure to Democrats becomes lower than Democrat exposure to Republicans. In these areas, many Republicans live with extremely low levels of exposure to Democrats and the typical Republican begins to experience the sharp partisan isolation characterizing the experience of the typical Democrat in more dense locations. However, isolation in rural areas is present not only for Republicans—modal Democrats in low-density places outside of urban areas, who are often rural Hispanics, African Americans and Native Americans, also have levels of exposure less than 0.10.

Even in the low-density outer suburbs of the major, and some minor, urban areas, a large proportion of Democrats have virtually no exposure to Republicans in their residential environment, and the modal Republican also has low levels of exposure (approximately 0.20) to Democrats. These low levels of exposure for both Republicans and Democrats in the low-density suburbs of major urban areas suggests that segregation is not solely a product of the sorting of Democrats into central cities and Republicans into suburbs—rather, even within the suburban fringes of urban areas, Democrats and Republicans are separated.

Partisans sort even within neighbourhoods. The severity of residential segregation demonstrates significant separation of partisans across the United States, but the extent of sorting can be better understood by comparing cross-party exposure of Republicans and Democrats living in the same places. For example, a proportion of the high isolation experienced by Democrats in the dense areas of large cities is to be expected, because many more Democrats than Republicans live in these large cities. But to what extent do Democrats and Republicans living in the same areas still have different levels of partisan exposure?

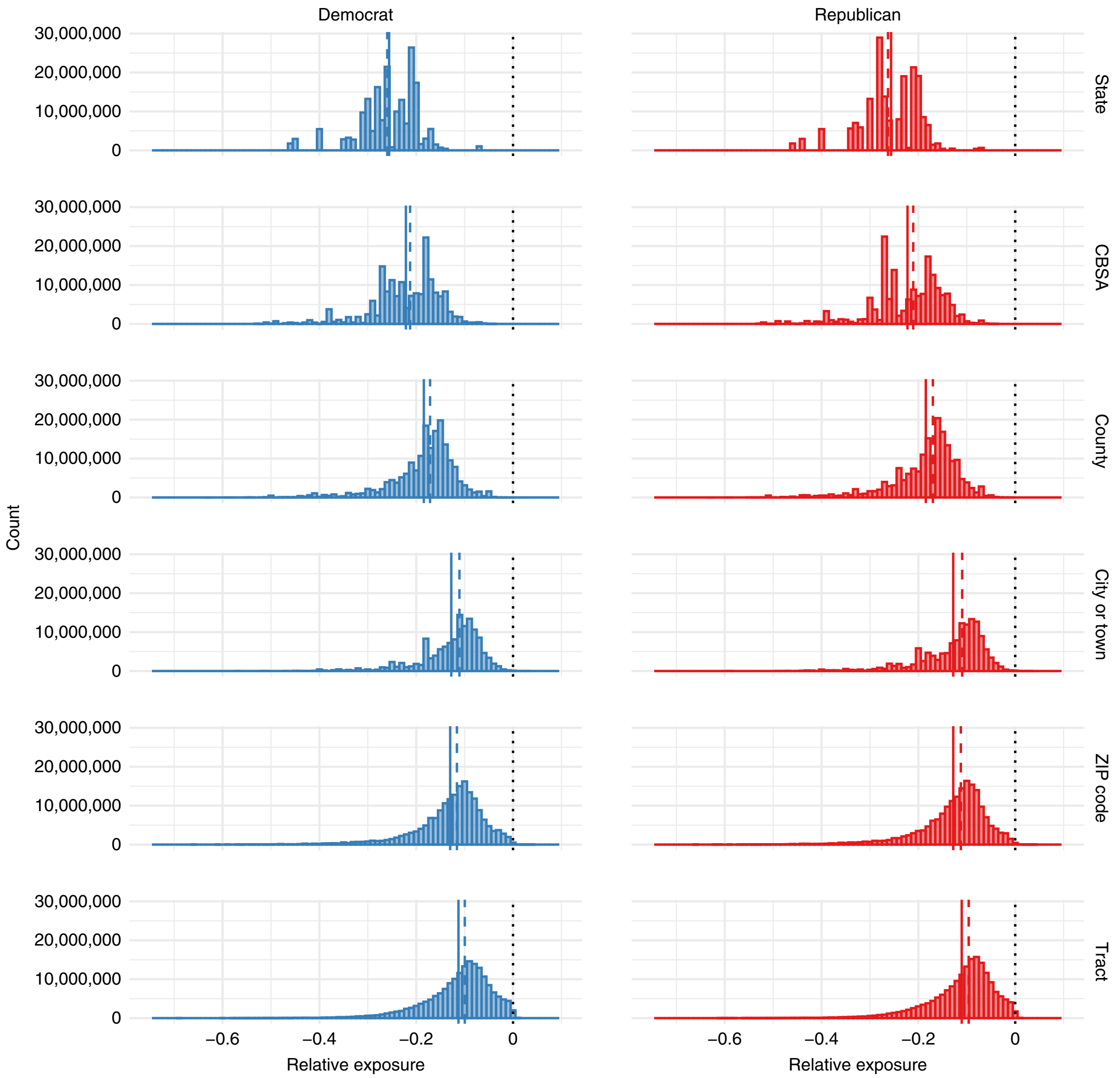


Fig. 5 | Relative exposure by geography. Histograms show the weighted nationwide distribution of relative exposure across geographic units for Democrats (blue) and Republicans (red). Distributions are weighted by population and the *y*-axis represents the number of individual voters. Solid vertical lines represent mean values and dashed lines represent median values. Geographies are ordered from bottom to top in decreasing average size. Sample size varies by coverage of the geographic variables (state, $n=180{,}660{,}182$; CBSA, $n=169{,}624{,}332$; county, $n=180{,}660{,}179$; city or town, $n=130{,}057{,}604$; ZIP code, $n=180{,}571{,}398$; census tract: $n=180{,}657{,}835$).

We compared the exposure of partisans to that of out-partisans living in the same geographic area, finding that even conditional on the choice to live in a larger geography, say within a particular city or even within the same neighbourhood, Democrats and Republicans still cluster with voters from their own party. We demonstrate this by constructing an index of relative exposure, defined as the difference between the average spatial exposure of partisans within a geographic unit to the average spatial isolation of out-partisans also living in that geographic unit. So, the relative exposure for Democrats in a geographic unit would be the difference between average Democrat exposure to Republicans and average Republican exposure to Republicans within that unit. This captures the extent to which members of one party experience different partisan environments than members of the other party living in the same geographic unit. If after moving into the geographic unit, no further sorting was occurring, we would expect this difference to be 0. For example, a city where Republicans have a relative exposure of −0.20 would be a city where out of a voter's 100 nearest neighbours, Republicans on average would have 20 more nearest neighbours who are also Republicans than the average Democrat has Republican neighbours. Notably, such comparisons across groups in the same geographic area are not possible when using traditional aspatial measures of exposure.

We calculate relative exposure across a range of geographies. Because we measure the exposure of individuals, we can measure relative exposure in any arbitrarily defined unit and therefore avoid problems such as the modifiable-areal-unit problem. However, to facilitate comparison with previous research, we use commonly used geographies and compare Democrats and Republicans living in the same state, urban area (CBSA), county, city or town, ZIP code and census tract.

At every level of geography, both Republicans and Democrats have relative exposure far lower than 0, indicating that substantial sorting does occur (Fig. 5). Large-scale sorting can be seen at the state level—probably driven by the clustering of Democrats into urban areas—but even within cities, Democrats and Republicans sort into different places, again indicating that partisan segregation is not merely a result of large-scale geographic trends such as an urban–rural divide. Moreover, even within census tracts, a small level of geography often used in social science research to represent

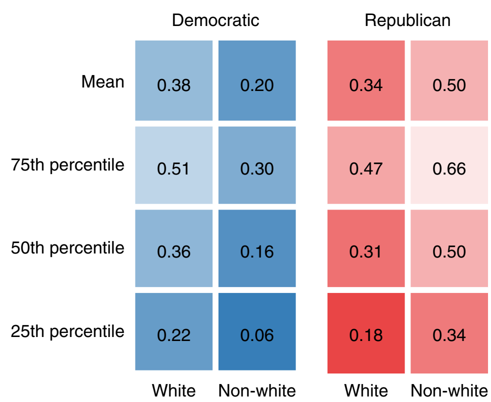


Fig. 6 | Nationwide percentiles of partisan exposure by party and race. Coloured cells show spatial exposure to out-party for non-Hispanic white (n = 115,736,045) and voters from all other racial and ethnic groups (n = 64,924,157) for Democrats (blue) and Republicans (red). Central tendencies and percentiles are from distributions weighted by posterior partisanship probabilities.

a neighbourhood, both Democrats and Republicans cluster more with their co-partisans than with out-partisans. Even at these small geographic levels, the difference between Democrats' exposure to Republicans and that of Republicans' to other Republicans (and vice versa) is 11 percentage points greater than we would expect if there was no partisan sorting within the geographic unit. This disparity indicates that, even after the Democrats and Republicans make the choice to live in similar neighbourhoods, they still live with noticeably different levels of partisan exposure (in the Supplementary Information, we demonstrate the robustness of these results to dropping neighbours who live in the same household as the voter).

Partisan segregation is distinct from racial or ethnic segregation. In the United States, race and ethnicity are highly correlated with partisanship, such that Republicans are much more likely than Democrats to be in the non-Hispanic white group. Much partisan segregation may be a function of the significant racial and ethnic segregation in the United States. Because we can observe the race and ethnicity of individuals on the voter file, we can examine how much partisan sorting is merely a function of racial/ethnic sorting.

We compare partisan exposure measured among all neighbours to the same measures but with exposure measured only among neighbours of the same race or ethnicity. Among non-Hispanic white voters, the distribution of the difference between exposure calculated among all voters and among only their non-Hispanic white neighbours is narrowly centred around 0, indicating that, on average, partisan exposure within race and ethnicity for non-Hispanic white voters mirrors general partisan segregation and, thus, partisan segregation is not only a function of racial or ethnic segregation (Extended Data Fig. 6 and Supplementary Information show distributions of the same test for voters from other racial and ethnic groups).

For non-Hispanic white Democrats, racial or ethnic segregation may actually reduce the levels of partisan segregation that would be expected if more non-Hispanic white Democrats lived near voters from other racial and ethnic groups. Partisan exposure, subset to non-Hispanic whites and voters from all other racial and ethnic groups by party (Figure 6), reveals high partisan isolation of Democrats who are members of racial and ethnic groups other than non-Hispanic white and, among the relatively small numbers of Republicans from racial and ethnic groups other than non-Hispanic white, the highest exposure of any group. This is probably attributable to the forces of racial and ethnic segregation clustering together voters who are not non-Hispanic whites, regardless of partisanship. Consequentially, the most isolated partisans are Democrats from racial and ethnic groups other than non-Hispanic white, indicating that the patterns of racial segregation may influence the creation of these clusters of extreme partisan isolation. By contrast, levels of exposure for non-Hispanic white Democrats and non-Hispanic white Republicans are largely similar to each other and higher than for Democrats from other racial or ethnic groups. This discrepancy between non-Hispanic white voters and voters from other racial or ethnic groups is suggestive evidence that non-Hispanic white voters sorting away from members of other racial or ethnic groups (for example, in ref. [51]) increases their exposure to non-Hispanic white members of the out-party, thus increasing the levels of partisan exposure over what would be present if partisan segregation were merely a function of racial or ethnic segregation.

Discussion

The isolation of voters by political party is a widely discussed feature of contemporary American politics. Uneven distributions of voters threatens equitable representation[14], and isolation from opposing political viewpoints may influence the development of partisan affect, policy preferences and patterns of political behaviour, and may also shape the strategies of campaigns and elected officials. Despite claims of a starkly segregated America, partisan segregation to date has been imprecisely measured, so the extent of partisan sorting was only understood across very large geographies.

By using geolocated records to develop a spatially weighted measure of exposure to neighbours of both parties for every voter in the United States, we move beyond conventional measures of segregation by measuring exposure at the individual level, circumventing common problems of aggregate measurement present in most measures of segregation, and by incorporating the distance between neighbours into our measurement. This yields an accounting of individual spatial segregation of partisans for the entire United States.

As we note throughout this article, our method involves measurement choices that may affect the conclusions drawn from our study. These include choices about the size of the set of neighbours used to construct indices of segregation; choices of weights on distance; and the imputation of partisanship. For each of these choices, we have shown the sensitivity of our results in the Supplementary Information. It is also important to note that we can only measure exposure to registered voters, so our measures cannot directly account for exposure to people who are not registered to vote. Further, although voter files are regularly cleaned to remove voters who have died, moved or otherwise left their location, some of these may remain in our analysis. We also measure distance using the most direct geodesic route between voters and do not account for features of the built environment, such as buildings and road networks, that may also affect exposure. Although we have no reason to suspect that these choices create systematic error, future researchers may want to account for these complexities. Finally, because partisan exposure may, in itself, cause changes in individual partisanship (J. R. Brown, unpublished manuscript), researchers should account for the dynamic features of partisan geography (J. R. Brown et al., unpublished manuscript) that our cross-sectional analysis does not capture.

Our results show high partisan segregation across the country, with most voters of both political parties living in partisan bubbles with little exposure to the other party. These high levels of isolation exist in different types of regions and at different population densities. Democrat exposure to Republicans is on average lower than Republican exposure to Democrats, markedly lower in high and medium population densities, and higher than Republican exposure to Democrats in low-density areas. Republican exposure is lowest in very-low-density areas. We also demonstrate that partisan segregation is distinct from racial and ethnic segregation, and that for non-Hispanic white voters, racial and ethnic sorting may reduce partisan isolation. Comparing Democrats and Republicans who live in the same city or neighbourhood, we find substantial differences in partisan environments, evidence that partisan sorting is driven by forces beyond the decision to live in specific, states, cities or neighbourhoods.

The high levels of partisan segregation can probably be explained by two broad and different, but not mutually exclusive, mechanisms that may provide fruitful areas for future research. First, previous research has shown that the correlation between residential density and partisanship has long historical roots related to Democrats and Republicans sorting into different types of housing, often as a function of occupation and income, producing stable geographic patterns of partisanship even across generations[41]. This points to partisan segregation arising largely from the immobility of voters. A second mechanism, consistent with the sorting that is found even within neighbourhoods and with the partisan sorting that is in addition to racial sorting, is the influence of micro-level behaviours on these large-scale patterns. While the best available evidence shows that most voters consider the partisan composition of an area to be low on their list of priorities when choosing neighbourhoods[52], it is still possible that partisan differences in income and lifestyle preferences, such as transportation and type of housing, may drive some voters to select different cities, neighbourhoods and, in some cases, streets or houses within neighbourhoods, even if partisanship is not an explicit criterion for selection. As partisanship becomes more correlated with lifestyle differences[53], such sorting may be further exacerbated.

Furthermore, there is evidence for party-based affective attitudes among Americans that are, by some measures, stronger than effects based on race[23]. Given that individual attitudes were responsible for some—although certainly not all—of the large-scale racial and ethnic sorting that occurred across neighbourhoods in the United States[1,34], it is possible that some voters, especially if they have already selected a city or neighbourhood in which to live, make decisions on the basis of the partisanship of their neighbours, driving some of the clustering that we observe. Moreover, if voters make decisions about where to live on the basis of the characteristics of their potential neighbours that are correlated with partisanship, such as their race or ethnicity and income, this can also drive the sorting we observe, both at large levels such as the city and small levels such as the neighbourhood.

Methods

This research was approved by the Harvard University Committee on the Use of Human Subjects. Informed consent was obtained from all survey participants. These participants did not receive monetary compensation for their participation.

Voter file. Our voter file was obtained from L2, a commercial data vendor working with both major political parties in the United States. The file contains 180,735,645 entries, reflecting the count of registered voters in the United States as of June 2018. We removed 75,443 entries could not be successfully geocoded, leaving 180,660,202. L2 and other commercial vendors obtain these data from state governments, who collect it for the purposes of administering elections and helping incumbent politicians in their campaigns[36]. The vendors then sell these data to political campaigns for voter targeting. Voter data provided by the states usually includes a large number of records that are invalid because the registrant has died or moved and re-registered. L2 and other vendors attempt to remove these records before selling the data. The commercial vendors also attempt to link people across time, preserving parts of their records as they change addresses and/or states delete older data.

Voters' race or ethnicity is available on voter files, either because it is recorded at the time of registration in some states or from imputation methods[46] similar to those we implemented for party in this study.

Data processing. The analysis in this manuscript relies on the spatial processing technique of geohashing. Geohash is a public domain encoding system that stores spatial location data in strings of letters and numbers. This technique is an application of z-order curves, transforming multi-dimensional data to a single dimension to allow for more efficient processing of such data[37]. These techniques were applied to the latitude and longitude coordinates of the residence of each registered voter in the United States, so that k-means analysis can identify the nearest neighbours and measure distances between neighbours for each voter in the nationwide voter file. This data-processing plan was developed in collaboration with Harvard University's Center for Geographic Analysis (CGA), which implemented the nearest-neighbours analysis and distance calculations. The processing was implemented on Amazon Web Services across ten Postgres instances with PostGIS add-ons. Processing of the entire file in this setup took approximately four weeks of continuous computation time. The output was a dataset of $1{,}000 \times 180{,}660{,}202$ rows listing the 1,000 nearest neighbours of each voter, with columns for the neighbour's partisanship and the distance they live from the voter.

With these data in hand, we calculated weighted averages of spatial partisan exposure and isolation and other partisan segregation, racial segregation and general summary statistics for each voter in the file. This stage of the analysis was also conducted on Amazon Web Services, using an instance calibrated for parallel processing in R. Total computation time for this stage was more than 200 h, completed in multiple instalments.

Table 1 | Density classifications by households per square mile

Classification	Households per square mile	Percent of Voters
Very low density	<102	22.92%
Low density	[102, 800)	28.90%
Medium density	[800, 2,213)	29.15%
High density	≥2,213	18.97%

We classify high, medium, low and very low density by the population per square mile of the census tract in which a voter lives, using the classifications developed by D. H. Montgomery of CityLab.

Partisanship imputation accuracy. We calculated the accuracy of our imputation of partisanship using a Brier score method to measure deviations of the imputed posterior partisan probabilities from the self-reported ideology of the survey respondents. This yields an accuracy of 77%. The error rate in this context does not mean that we incorrectly predict 23% of voters' partisan exposure. Since our measures of spatial partisan exposure and isolation are weighted averages of the probability of republican and democratic affiliation across all neighbours, when our probabilistic forecast underestimates the extent to which a voter prefers the Republican to the Democratic party, or vice versa, we are directly accounting for this uncertainty (Supplementary Information).

Examining other data from our survey and previous panel surveys gives a sense of the expected upper bound of accuracy in surveying partisan identity, which our 77% accuracy approaches. Our survey included registered voters and the match between registered party and self-reported partisan identity from the survey is 84% (84.31% for Republicans and 83.37% for Democrats) (Extended Data Fig. 5). This is similar to what we might expect on the basis of temporal instability of partisan identification in panel surveys[19], which report stability of self-reported partisanship in panel surveys of 0.80 to 0.85 over 2 years, 0.82 to 0.84 over 1 year, and 0.83 to 0.90 over less than 1 year. In most cases with our data, the registration happened many years before the survey, so temporal stability of 84% slightly exceeds what might be expected from the two-year findings of ref. [19]. Furthermore, assuming that partisanship as a social identity is unlikely to change over a short time period, the instability of 0.10 to 0.17 in a period of less than a year may be largely attributable to survey error and, therefore, gives an estimate of survey instability that comes from stochastic fluctuations or survey error. This implies an expected upper bound of a match between a survey response and another measure of partisanship of 0.83 to 0.90.

We can also use our imputed scores to assign each voter a discrete party affiliation based on the party with the highest probability from the imputation. Through this process, we classify 89% of voters not registered to a major party as leaning toward either Democrats or Republicans, which is very close to the proportion of non-partisan voters who consistently vote either Democrat or Republican according to previous studies[43]. With party assigned in a discrete manner, our measures are not weighted by party. In the Supplementary Information, we present results with party classified in this way.

Nationwide proportion of Democrats and Republicans. The nationwide proportion of Democrats and Republicans of 51% and 43%, respectively, is calculated by taking the average of posterior partisan probabilities across all US voters.

Density classifications. We classify high, medium, low and very low density by the population per square mile of the Census Tract in which a voter lives, using the classifications developed by D. H. Montgomery of CityLab (Table 1).

Table 2 | Number of registered voters in quantiles by geography by geography

Geography	0.1%	1%	10%	25%	50%	75%	90%	99%	99.9%	Mean
State	237,298	283,212	478,367	976,729	2,798,419	4,334,925	7,643,908	15,710,329	17,948,081	3,542,357
CBSA	6,377	7,637	14,757	22,027	43,011	106,995	338,680	2,885,697	6,720,367	184,977
County	179	628	2,794	6,173	14,734	38,272	120,097	763,377	1,892,687	57,480
City/Town	2	10	68	183	622	2,456	8,595	55,440	295,080	4,542
Zip Code	1	2	94	353	1,460	7,114	16,900	33,084	45,909	5,272
Tracts	4	248	1,082	1,580	2,275	3,149	4,120	6,660	11,578	2,489

Geographic definitions and sources. We use geographic units of state, urban area, county, city or town, ZIP code, census tract and precincts. Precincts are used to create the imputations of partisanship. All geographic units except precincts are from the United States Census definitions from 2010 or later. Urban areas are defined by CBSAs and city and town are defined by census places, which includes incorporated places and census-designated places that are not incorporated.

Voters' state, county, ZIP code and census tract were taken directly from the voter file. Urban area is defined by county location extracted from the voter file and city or town is defined by a spatial merge of voters individual locations' with shapefiles provided by the Census Bureau. Precincts were collected from the MIT Election Lab (https://electionlab.mit.edu/) and the authors' own searches of state election administrations.

The distribution of registered voters in each geographic unit is listed in Table 2 and the percentage of voters with non-missing cases for each of these variables is listed in the Supplementary Information.

Relative exposure. Relative exposure is the difference between the average spatial exposure of partisans within a geographic unit to the average spatial isolation of out-partisans also living in that geographic unit:

$$\text{Relative exposure}_{g,p} = \frac{\sum_{i\in\{g\}} P(p)_i \text{Exposure}_{i,p}}{\sum_{i\in\{g\}} P(p)_i} - \frac{\sum_{i\in\{g\}} P(q)_i \text{Isolation}_{i,q}}{\sum_{i\in\{g\}} P(q)_i}$$

where relative exposure for party p within geographic unit g is the weighted average (weighted by, $P(p)_i$, the posterior probability of voter i being in-party p), of party p's exposure to out-party q, across all voters i in geographic unit g, minus the weighted average (weighted by $P(q)_i$) of party q's isolation, or exposure to itself.

We compare Democrats and Republicans living in the same state, urban area, county, city or town, ZIP code and census tract. At each level of geography, for both Democrats and Republicans, a population-weighted t-test for a mean different than zero yields $P<0.001$ (two-tailed test). Further details are provided in the Supplementary Information.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Anonymized replication data are available in the Harvard University Dataverse at https://doi.org/10.7910/DVN/A40X5L.

Code availability

All replication code are available in the Harvard University Dataverse at https://doi.org/10.7910/DVN/A40X5L.

Received: 9 June 2020; Accepted: 2 February 2021;
Published online: 08 March 2021

References

1. Massey, D. S. & Denton, N. A. *American Apartheid: Segregation and the Making of the Underclass* (Havard Univ. Press, 1993).
2. Alesina, A., Baqir, R. & Easterly, W. Public goods and ethnic divisions. *Q. J. Econ.* **114**, 1243–1284 (1999).
3. Trounstine, J. Segregation and inequality in public goods. *Am. J. Political Sci.* **60**, 709–725 (2016).
4. Enos, R. D. *The Space Between Us: Social Geography and Politics* (Cambridge Univ. Press, 2017).
5. Allport, G. W. *The Nature of Prejudice* (Addison–Wesley, 1954).
6. Pettigrew, T. F. & Tropp, L. A meta-analytic test of intergroup contact theory. *J. Personal. Soc. Psychol.* **90**, 751–783 (2006).
7. Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
8. Putnam, R. D. *Bowling Alone: The Collapse and Revival of American Community* (Simon & Schuster, 2001).
9. Fowler, J. H. & Christakis, N. A. Cooperative behavior cascades in human social networks. *Proc. Natl Acad. Sci. USA* **107**, 5334–5338 (2010).
10. Rand, D. G., Arbesman, S. & Christakis, N. A. Dynamic social networks promote cooperation in experiments with humans. *Proc. Natl Acad. Sci. USA* **108**, 19193–19198 (2011).
11. Uslaner, E. M. *Segregation and Mistrust: Diversity, Isolation, and Social Cohesion* (Cambridge Univ. Press, 2012).
12. Ananat, E. O. The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *Am. Econ. J. Appl. Econ.* **3**, 34–66 (2011).
13. Ananat, E. O. & Washington, E. Segregation and black political efficacy. *J. Public Econ.* **93**, 807–822 (2009).
14. Chen, J. & Rodden, J. Unintentional gerrymandering: political geography and electoral bias in legislatures. *Q. J. Political Sci.* **8**, 239–269 (2013).
15. Bishop, B. *The Big Sort* (Houghton Mifflin Harcourt, 2009).
16. Abrams, S. J. & Fiorina, M. P. The big sort that wasn't: a skeptical reexamination. *PS Political Sci. Politics* **45**, 203–210 (2012).
17. Nall, C. The political consequences of spatial policies: How interstate highways facilitated geographic polarization. *J. Politics* **77**, 394–406 (2015).
18. Martin, G. J. & Webster, S. W. Does residential sorting explain geographic polarization? *Political Sci. Res. Methods* **8**, 215–231 (2020).
19. Green, D. P., Palmquist, B. & Schickler, E. *Partisan Hearts and Minds* (Yale Univ. Press, 2004).
20. Achen, C. H. & Bartels, L. M. *Democracy for Realists: Why Elections do not Produce Responsive Government* (Princeton Univ. Press, 2017).
21. Hersh, E. D. & Goldenberg, M. N. Democratic and republican physicians provide different care on politicized health issues. *Proc. Natl Acad. Sci. USA* **113**, 11811–11816 (2016).
22. Chen, M. K. & Rohla, R. The effect of partisanship and political advertising on close family ties. *Science* **360**, 1020–1024 (2018).
23. Iyengar, S. & Westwood, S. J. Fear and loathing across party lines: new evidence on group polarization. *Am. J. Political Sci.* **59**, 690–707 (2015).
24. Aarts, H. & Dijksterhuis, A. The silence of the library: environment, situational norm, and social behavior. *J. Personal. Soc. Psychol.* **84**, 18–28 (2003).
25. Cialdini, R. B., Reno, R. R. & Kallgren, C. A. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *J. Personal. Soc. Psychol.* **58**, 1015–1026 (1990).
26. Green, D. P. et al. The effects of lawn signs on vote outcomes: results from four randomized field experiments. *Elect. Stud.* **41**, 143–150 (2016).
27. Bonica, A. Mapping the ideological marketplace. *Am. J. Political Sci.* **58**, 367–386 (2014).
28. Huckfeldt, R. & Sprague, J. Networks in context: the social flow of political information. *Am. Political Sci. Rev.* **81**, 1197–1216 (1987).
29. Zaller, J. R. *The Nature and Origins of Mass Opinion* (Cambridge Univ. Press, 1992).
30. Converse, P. E. in *Ideology and Discontent* (ed. Apter, D. E.) 206–261 (Free Press, 1964).
31. Verba, S., Schlozman, K. L. & Brady, H. E. *Voice and Equality: Civic Voluntarism in American Politics* (Harvard Univ. Press, 1995).
32. Enos, R. D. & Gidron, N. Intergroup behavioral strategies as contextually determined: experimental evidence from Israel. *J. Politics* **78**, 851–867 (2016).
33. Cramer, K. J. *The Politics of Resentment: Rural Consciousness in Wisconsin and the Rise of Scott Walker* (Univ. Chicago Press, 2016).
34. Schelling, T. C. *Micromotives and Macrobehavior*. (Norton, 1971).
35. Enos, R. D. & Hersh, E. D. Party activists as campaign advertisers: the ground campaign as a principal-agent problem. *Am. Political Sci. Rev.* **109**, 252–278 (2015).
36. Hersh, E. *Hacking the Electorate: How Campaigns Perceive Voters*. (Cambridge Univ. Press, 2015).
37. Niemeyer, G. Tips & tricks. *Geohash.org* http://geohash.org/site/tips.html (2008).
38. Reardon, S. F. & O'Sullivan, D. Measures of spatial segregation. *Sociological Methodol.* **34**, 121–162 (2004).

39. Massey, D. S. & Denton, N. A. The dimensions of residential segregation. *Soc. Forces* **67**, 281–315 (1988).
40. White, M. J. The measurement of spatial segregation. *Am. J. Sociol.* **88**, 1008–1018 (1983).
41. Rodden, J. A.*Why Cities Lose: The Deep Roots of the Urban–Rural Political Divide* (Basic Books, 2019).
42. Keith, B. E. et al. *The Myth of the Independent Voter* (Univ. California Press, 1992).
43. Magleby, D., Nelson, C. & Westlye, M. in *Facing the Challenge of Democracy: Explorations in the Analysis of Public Opinion and Political Participation* (eds Sniderman, P. M. & Hifhton, B.) 238–263 (Princeton Univ. Press, 2011).
44. Hawkins, C. B. & Nosek, B. A. Motivated independence? Implicit party identity predicts political judgments among self-proclaimed independents. *Personal. Soc. Psychol. Bull.* **38**, 1437–1452 (2012).
45. Klar, S. & Krupnikov, Y. *Independent Politics* (Cambridge Univ. Press, 2016).
46. Enos, R. D. What the demolition of public housing teaches us about the impact of racial threat on political behavior. *Am. J. Political Sci.* **60**, 123–142 (2016).
47. Imai, K. & Khanna, K. Improving ecological inference by predicting individual ethnicity from voter registration records. *Political Anal.* **24**, 263–272 (2016).
48. Converse, P. E. in *Elections and the Political Order* (eds Campbell, A., et al.) 9–39 (Wiley, 1966).
49. Ansolabehere, S. & Snyder, J. M. Jr The incumbency advantage in US elections: an analysis of state and federal offices, 1942–2000. *Elect. Law J.* **1**, 315–338 (2002).
50. Oakes, P. in *Rediscovering the Social Group: A Self-categorization Theory* (eds Turner, J. C. et al.) 117–141 (Blackwell, 1987).
51. Farley, R., Steeh, C., Krysan, M., Jackson, T. & Reeves, K. Stereotypes and segregation: neighborhoods in the Detroit area. *Am. J. Sociol.* **100**, 750–780 (1994).
52. Mummolo, J. & Nall, C. Why partisans do not sort: the constraints on political segregation. *J. Politics* **79**, 45–59 (2017).
53. Hetherington, M. & Weiler, J. *Prius Or Pickup? How the Answers to Four Simple Questions Explain America's Great Divide* (Houghton Mifflin, 2018).

Acknowledgements

The authors received no specific funding for this work. We thank B. Lewis and D. Kakkar at the Harvard Center for Geographic Analysis, the Harvard MIT Data Center, and A. Dagonel for research assistance; M. Schwenzfeier and J. Rodden for advice on research design; A. Agadjanian for help making sense of the precinct data; N. Cohn for providing survey data on the age distribution of voters in Wisconsin; and seminar participants at the University of Pittsburgh, New York University, Northeastern University, Princeton University, Brown University, University of Massachusetts at Amherst and Harvard University.

Author contributions

J.R.B. and R.D.E. both contributed to the conception, design, analysis, data collection, and writing.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-021-01066-z.

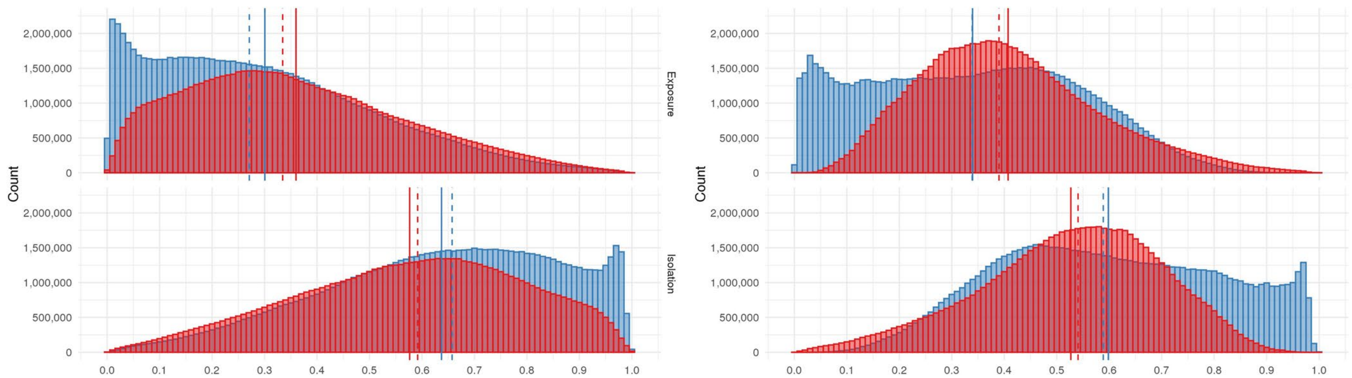
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-021-01066-z.

Correspondence and requests for materials should be addressed to J.R.B. or R.D.E.

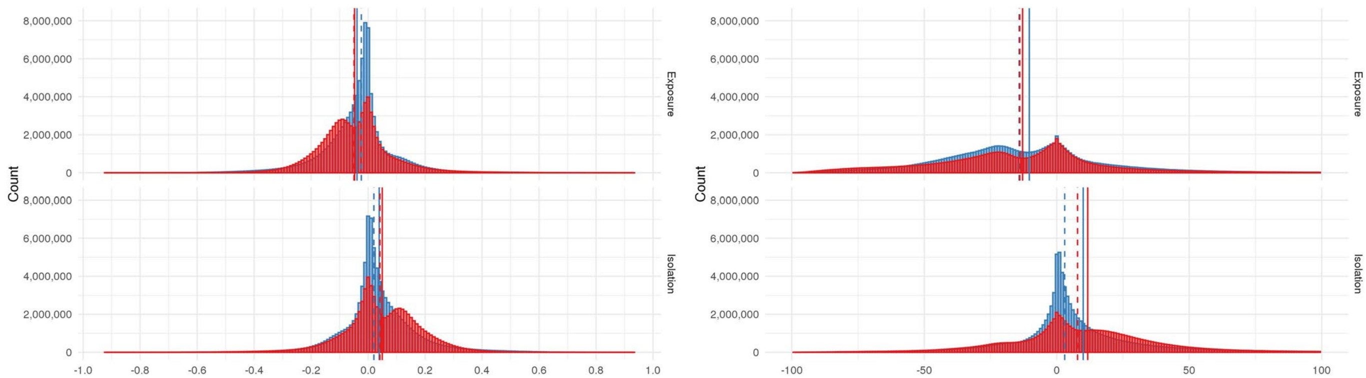
Peer review information *Nature Human Behaviour* thanks M. Keith Chen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Aisha Bradshaw.

Reprints and permissions information is available at www.nature.com/reprints.

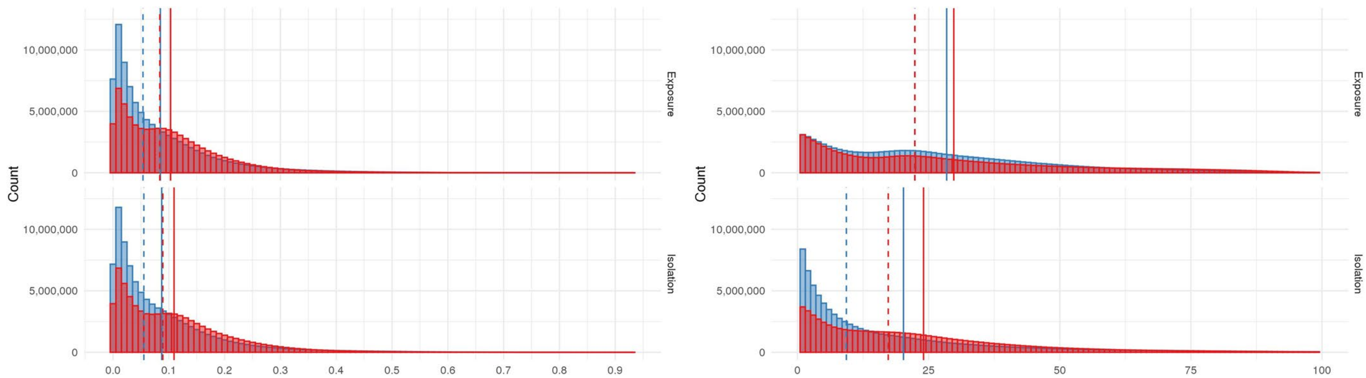
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



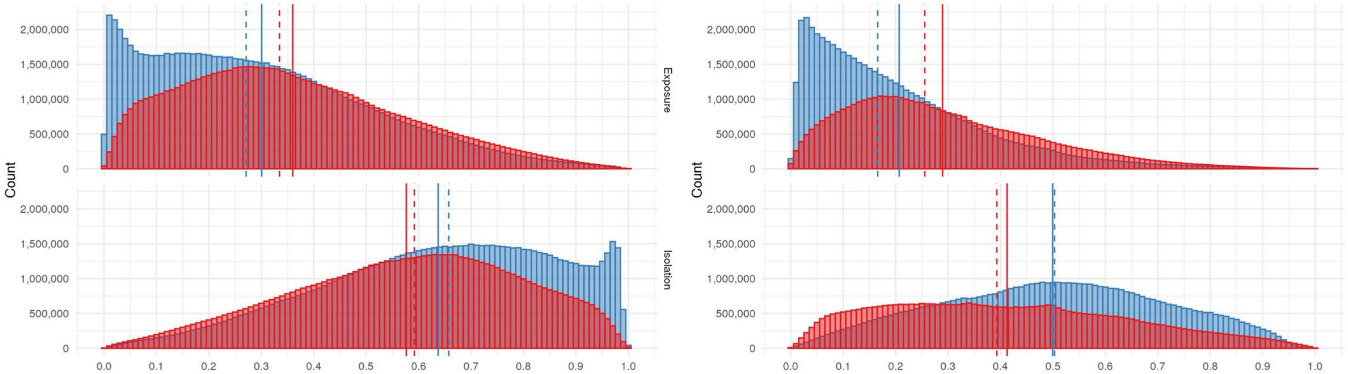
Extended Data Fig. 1 | Spatial versus Aspatial Exposure/Isolation. Nationwide distribution (n = 180,660,202) of individual spatial (left) and aspatial (right) partisan isolation and exposure separately for Democrats (blue) and Republicans (red). Solid vertical lines represent mean values and dashed lines represent median values. The distributions are weighted by the posterior partisan probabilities.



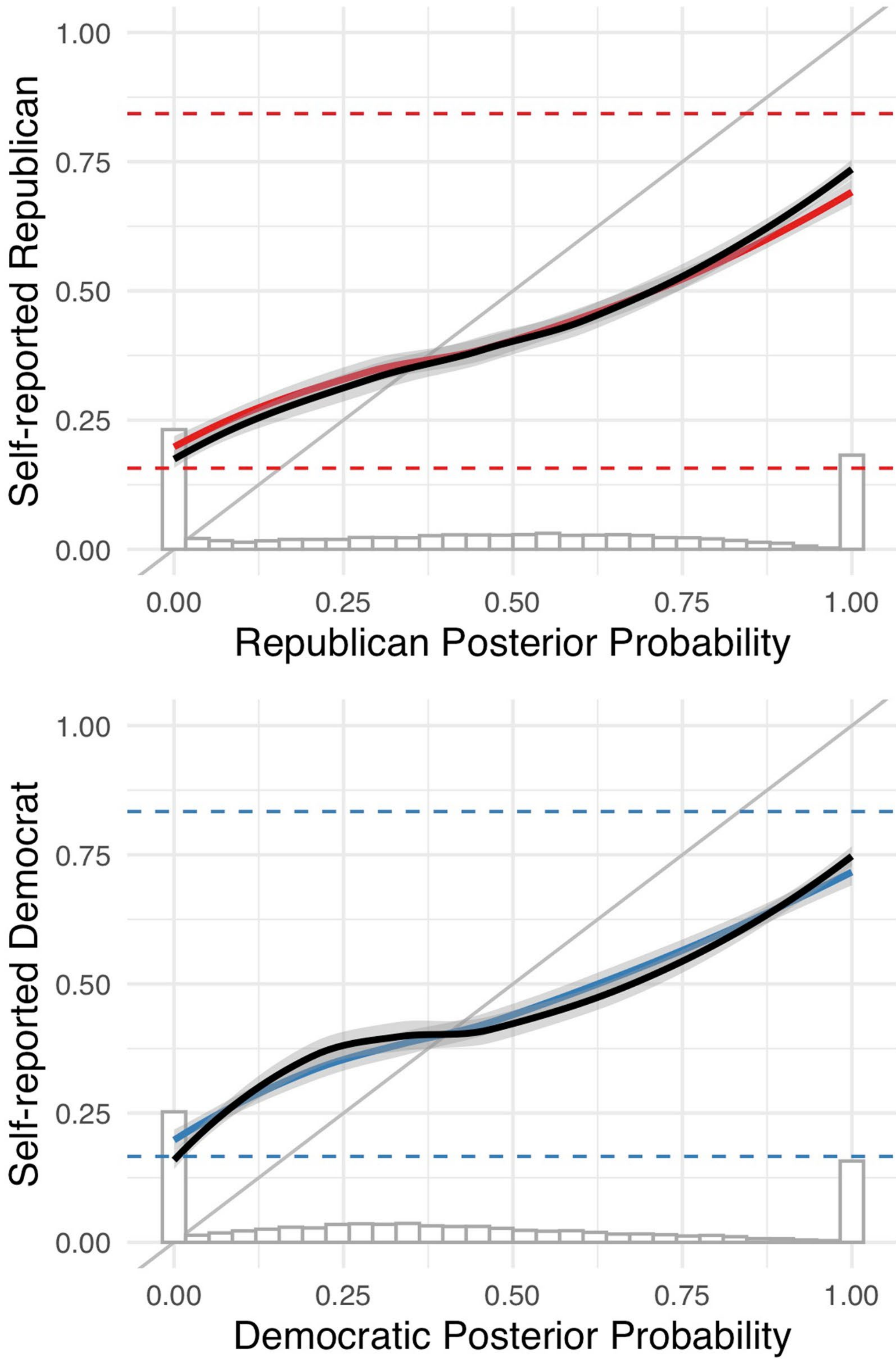
Extended Data Fig. 2 | Individual Differences in Spatial versus Aspatial Exposure/Isolation. Nationwide distribution (n = 180,660,202) of individual-level changes in partisan Exposure and Isolation separately for Democrats (blue) and Republicans (red). The histograms on the left show the percentage point difference in spatial and aspatial exposure, while the histograms on the right show the percent change. Solid vertical lines represent mean values and dashed lines represent median values. The distributions are weighted by the posterior partisan probabilities.



Extended Data Fig. 3 | Individual Absolute Differences in Spatial versus Aspatial Exposure/Isolation. Nationwide distribution (n = 180,660,202) of individual-level absolute changes in partisan Exposure and Isolation separately for Democrats (blue) and Republicans (red). The histograms on the left show the percentage point absolute difference in spatial and aspatial exposure, while the histograms on the right show the absolute percent change. Solid vertical lines represent mean values and dashed lines represent median values. The distributions are weighted by the posterior partisan probabilities.

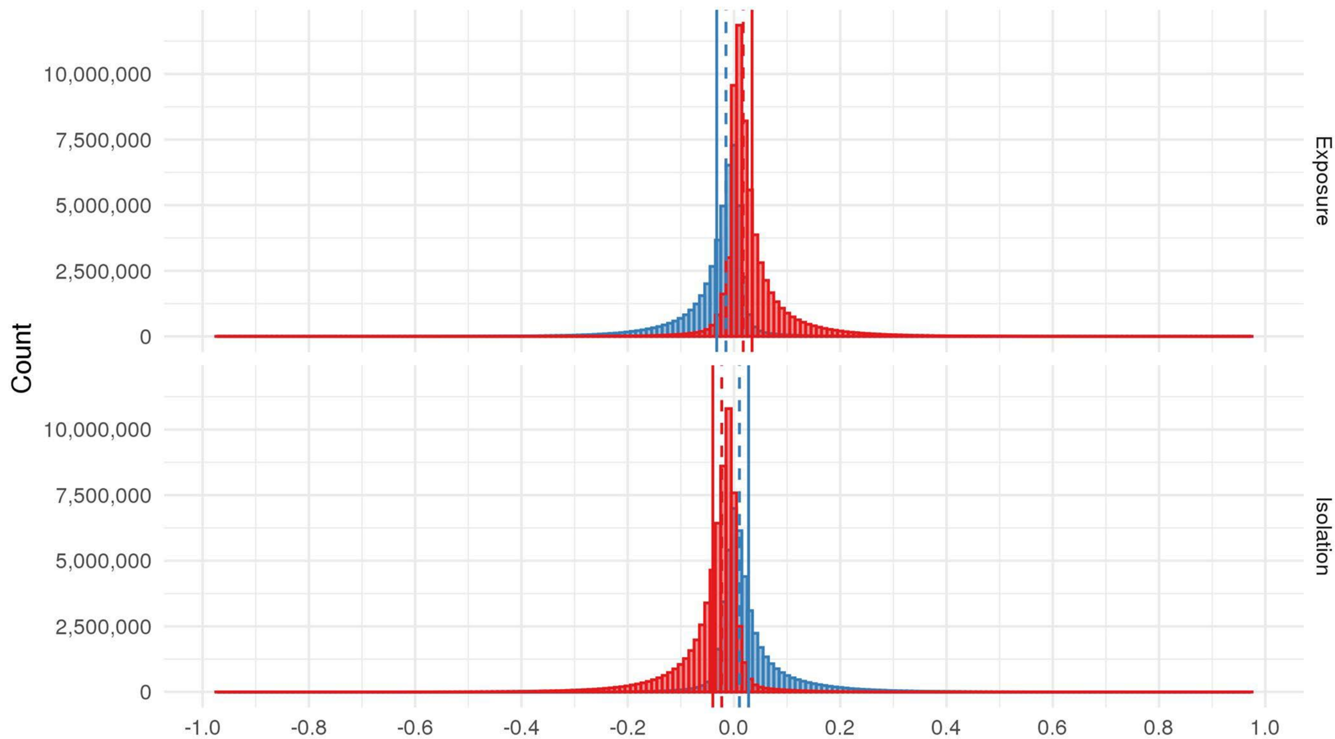


Extended Data Fig. 4 | Exposure and Isolation with Imputation Versus Without Imputation. Nationwide distribution (n = 180,660,202) of individual spatial partisan isolation and exposure with imputation of partisanship (left) and without (right) separately for Democrats (blue) and Republicans (red). Solid vertical lines represent mean values and dashed lines represent median values. The distribution on the left is weighted by the posterior partisan probabilities.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Percent self-report Partisan Category by Posterior Partisan Probability. LOESS lines plotting the relationship between posterior partisan probability (Republicans on top, Democrats on bottom) and the rates of survey respondents reporting as the corresponding partisanship. The correlation is limited to the subset of survey respondents ($n = 7,087$) who are not registered with a major political party. Black lines plot the LOESS curve with survey weights incorporated, red/blue lines without survey weights. The 45-degree grey line plots a perfect 1-to-1 relationship between posterior partisan probability and self-reported partisanship. The horizontal dotted lines show the rates at which survey respondents who are registered Democrats/Republicans self-report partisanship in agreement (or disagreement for the lower lines) with their actual partisan registration. That is, the upper blue (red) dotted line represents the proportion of survey respondents we know are registered Democrats (Republicans) who self-report as Democrats (Republicans), and the lower dotted line represents the proportion who do not self-report as Democrats (Republicans). These lines represent lower and upper bounds on how accurate we can expect our forecast to appear when measured against survey data. The histogram on the bottom plots the frequency distribution of posterior partisan probabilities across the unaffiliated subset.



Extended Data Fig. 6 | Partisan Segregation vs. non-Hispanic White-only Partisan Segregation. Distribution for non-Hispanic white voters (n = 115,736,045) of differences between partisan segregation calculated from all 1,000 nearest neighbors and partisan segregation calculated only from non-Hispanic white neighbors. Positive Isolation values means that a voter appears less isolated by partisanship when we look only at their non-Hispanic white neighbors. Positive Exposure values means that a voter appears to have less cross-party exposure when we only look at their white neighbors. Distributions are plotted separately for Democrats (blue) and Republicans (red). Solid lines represent mean values and dashed lines represent median values. Distributions are weighted by posterior partisan probabilities.

nature research

Corresponding author(s): Ryan Enos

Last updated by author(s): 12/9/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect data for this project.
Data analysis	Identification of closest neighbors and distance calculation was conducted in PostGIS. Analysis of these data and production of all figures in the paper were conducted in R. Maps were produced using R, except for nearest neighbor maps, which were produced in ArcGIS.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All replication materials are available in the Harvard University Dataverse: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/A40X5L

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative observational data
Research sample	The sample of study is the population of U.S. registered voters in 2018. Data are collected from nationwide voterfiles listing every registered voter in the country as of the date of the voterfiles. We choose this sample because we this project analyzes partisan sorting for every voter using individual data, and thus the entirety of the voting population is necessary for this fine-grained analysis. Supplementary analyses use an original survey of 12,221 registered voters. This supplementary analysis is used to validate imputations of partisanship, and we survey a sample of voters in our larger study to directly validate imputations, connecting self-reported partisanship and ideology to our individual imputations.
Sampling strategy	The sample is defined as all U.S. registered voters listed in nationwide voterfiles as of 2018. Inclusion in the study hinges on a voter being recorded in these voterfiles. Potential survey respondents were randomly sampled after stratification by state and whether partisanship was visible on the voterfile, with an over-sample of non-partisans. Respondents were contacted by email from email addresses linked to the voterfile by the vendor L2. We sent emails to 1,753,493 unique voters. Of these e-mails, 47.2% bounced, indicating that the e-mail was invalid or that our e-mail was rejected by the server, perhaps for spam protection. Thus, 925,339 unique voters received an invitation to participate in our survey, and we received 12,221 responses, as response rate of 1.3%, which is similar to the single-digit response rates expected for modern phone or e-mail surveys.
Data collection	Voterfile data was acquired from the vendor L2, a non-partisan firm that provides voterfile data to campaigns and researchers. Survey data were collected by email from email addresses linked to the voterfile by the vendor L2. Surveys were administered by Qualtrics and were accessed by consenting respondents through an e-mail link.
Timing	All voterfiles for our study are current as of June 2018. Survey data for supplementary analyses were collected between June 10, 2019 and October 30, 2019.
Data exclusions	Of the 180,735,645 registered US voters in our study, 75,443 were removed from the final analyses because they could not be successfully geocoded.
Non-participation	Four our survey, we sent emails to 1,753,493 unique voters. Of these e-mails, 47.2% bounced, indicating that the e-mail was invalid or that our email was rejected by server, perhaps for spam protection. Thus 925,339 unique voters received an invitation to participate in our survey and we received 12,221 responses for a response rate of 1.3%
Randomization	Participants were not allocated into randomized groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Supplementary analyses use an original survey of 12,221 registered voters. Potential survey respondents were randomly sampled after stratification by state and whether partisanship was visible on the voterfile, with an over-sample of non-partisans. Population characteristics reflect characteristics of the U.S. registered voters, and thus cannot be below the age of 18, and must be registered to vote.
Recruitment	Potential survey respondents were randomly sampled after stratification by state and whether partisanship was visible on the voterfile, with an over-sample of non-partisans. Respondents were contacted by email from email addresses linked to the voterfile by the vendor L2. We sent emails to 1,753,493 unique voters. Of these e-mails, 47.2% bounced, indicating that the e-mail was invalid or that our email was rejected by server, perhaps for spam protection. Thus 925,339 unique voters received an invitation to participate in our survey and we received 12,221 responses for a response rate of 1.3%, which is similar to the single digit response rate for modern phone or email surveys.
Ethics oversight	Harvard IRB

Note that full information on the approval of the study protocol must also be provided in the manuscript.